

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/320009252>

# Autonomia delle macchine e filosofia dell'intelligenza artificiale

Article *in* Rivista di filosofia · August 2017

CITATIONS

0

READS

43

1 author:



Guglielmo Tamburrini

University of Naples Federico II

83 PUBLICATIONS 430 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



REFILLS H2020 - Ethics Work Package [View project](#)



Computational models of neuromodulation in the control of plastic behaviours [View project](#)

Penultimate version of the article “Autonomia delle macchine e filosofia dell’intelligenza artificiale”, in *Rivista di filosofia* **58**(2), 2017, pp. 263-275.

Rivista di filosofia : <https://www.mulino.it/riviste/issn/0035-6239>

## **Autonomia delle macchine e filosofia dell’intelligenza artificiale**

GUGLIELMO TAMBURRINI<sup>1</sup>

Dipartimento di Ingegneria Elettrica e Tecnologie dell’Informazione, Università di Napoli Federico II  
[guglielmo.tamburrini@unina.it](mailto:guglielmo.tamburrini@unina.it)

### *1. Autonomia delle macchine: centralità della questione etica*

Fin dalla nascita dell’intelligenza artificiale (IA) intorno alla metà del secolo scorso, i filosofi si sono diffusamente interrogati sulla questione dell’intelligenza delle macchine e delle sue relazioni con l’intelligenza umana. A questa problematica – esplorata dalle prospettive della filosofia della mente, della scienza e della tecnologia – si è affiancata più di recente l’esigenza di sviluppare una riflessione sulla *autonomia* delle macchine.

Le tecnologie dell’IA consentono di costruire macchine che sostituiscono gli esseri umani in una varietà sempre crescente di compiti: veicoli senza conducente, algoritmi che eseguono transazioni finanziarie più velocemente di quanto un essere umano possa mai fare, sistemi d’arma che identificano e attaccano i loro bersagli senza essere preventivamente sottoposti all’autorizzazione di un operatore umano, sistemi per la diagnosi medica, agenti autonomi che penetrano nei sistemi informatici del nemico o proteggono da analoghi attacchi sferrati nel contesto della guerra cibernetica.

In ciascuno di questi casi, una macchina può fare in autonomia qualcosa che comporterebbe l’assunzione di responsabilità morali da parte di un esecutore umano. Per questa ragione, l’interesse filosofico per i sistemi autonomi è in buona misura alimentato da una questione etica: in quali circostanze sia moralmente permesso o proibito far svolgere autonomamente a una

---

<sup>1</sup> L’autore ringrazia Daniele Amoroso per stimolanti discussioni sui temi trattati. Le relative attività di ricerca sono state svolte con il sostegno economico del MIUR (progetto PRIN 2015, cod. 2015TM24JS\_009).

macchina qualcosa che ha implicazioni significative sul piano delle responsabilità, dei doveri morali e dei diritti fondamentali delle persone.

Casi particolari della questione etica riguardano i veicoli senza conducente, in considerazione degli incidenti stradali nei quali essi possono essere coinvolti, gli squilibri economici che possono essere provocati dagli algoritmi autonomi per giocare in borsa, la soppressione di vite umane da parte di un'arma autonoma, le conseguenze di una diagnosi errata formulata da una macchina, i danni procurati alle popolazioni civili dagli agenti autonomi che in una guerra cibernetica paralizzano le principali infrastrutture di servizio di un intero territorio.

La riflessione etica sulle macchine autonome richiede di contestualizzare una pluralità di teorie etiche normative, individuando gli eventuali conflitti tra gli atteggiamenti morali che ciascuna di esse suggerisce in relazione alla progettazione e all'uso delle macchine autonome. In caso di conflitto interteorico, si pone l'ulteriore problema di indagare se vi sia una composizione coerente del conflitto che consenta di individuare una piattaforma etica ampiamente condivisa per guidare la progettazione e l'utilizzazione responsabile delle macchine autonome.

Questa articolazione schematica della questione etica a proposito dei sistemi autonomi sarà qui esplorata in relazione ai veicoli senza conducente e alle armi autonome. La questione etica che ci accingiamo ad esplorare in questo modo presuppone che si abbia un'idea relativamente chiara di che cosa sia una macchina autonoma. Di questo problema preliminare ci occuperemo brevemente nel prossimo paragrafo.

## 2. Che cos'è una macchina autonoma?

Varie impostazioni sono state date al problema di chiarire cosa sia una macchina autonoma. L'attenzione si è focalizzata di volta in volta su *i*) i compiti che la macchina deve svolgere senza interventi esterni da parte di altri agenti naturali o artificiali, *ii*) le capacità percettive, cognitive e decisionali delle quali essa deve essere dotata, *iii*) la sua architettura e organizzazione funzionale. E' stato perseguita anche la strada di considerare congiuntamente gli aspetti *i*)-*iii*)<sup>2</sup>.

---

<sup>2</sup> Per quest'ultima impostazione e per interessanti osservazioni sulle altre impostazioni al problema dell'autonomia, si veda il saggio di Giovanni Sartor e Andrea Omicini, *The autonomy of technological systems and responsibilities for their use*, in *Autonomous Weapons Systems: Law, Ethics, Policy*, a cura di Nehal Bhuta et al., Cambridge UP, 2016, pp. 39-74.

L'alternativa *i)* privilegia i compiti da svolgere senza l'intervento degli esseri umani rispetto a un insieme di proprietà – variamente articolate in base a *ii)-iii)* in termini di capacità, processi e organizzazione interna – delle quali deve godere la macchina in quanto condizioni necessarie per l'autonomia. La prima alternativa è perseguibile in relazione ai domini d'indagine per l'IA nei quali si può separare la descrizione di che cosa un sistema dell'IA deve fare da una considerazione delle varie capacità che risultano essere necessarie o sufficienti per farlo – a partire dal riconoscimento percettivo di varie classi di oggetti (ad esempio un volto umano o un'automobile di un certo modello), fino a problemi di manipolazione e di navigazione robotica, come pure alla descrizione delle regole di giochi individuali o di squadra (scacchi, calcio e così via).

Supponiamo che un dato compito *C* sia così descrivibile. Perché un sistema *S* possa dirsi autonomo nello svolgimento di *C*, non basta che *S* abbia svolto *C* in qualche occasione. Si richiede piuttosto che *S* svolga *regolarmente* *C* senza l'intervento di un essere umano. Un sistema formato da una caldaia e da un termostato che ne controlla accensione e spegnimento è autonomo in questo senso, poiché riesce a mantenere nel tempo la temperatura di un dato ambiente vicina a un valore prefissato senza interventi da parte di un essere umano. Perché questa regolarità sia soddisfatta devono naturalmente valere le giuste condizioni iniziali e al contorno. Condizioni al contorno di particolare rilevanza nel caso dei sistemi ingegneristici riguardano il *dove* dello svolgimento: il sistema costituito da termostato e caldaia non è stato né progettato né testato per svolgere regolarmente il suo compito in un appartamento con finestre spalancate e temperatura esterna molto bassa.

Si consideri ora il compito dei veicoli senza conducente: esso consiste, in prima approssimazione, nel circolare su una parte o su tutta la rete viaria per i veicoli con conducente. Un ulteriore vincolo riguarda il livello prestazionale del sistema di controllo dei veicoli senza conducente, che deve essere almeno altrettanto buono di quello di un conducente umano competente. Si tratta di una richiesta di carattere comparativo che è stata spesso avanzata per delimitare gli obiettivi dell'IA e apprezzarne i risultati. Basta menzionare qui il gioco dell'imitazione ideato da Alan Turing, l'obiettivo pienamente raggiunto di sviluppare un programma che giochi a scacchi al livello di gran maestro internazionale, l'obiettivo di lungo periodo dei tornei *Robocup* di costruire squadre di robot in grado di battere squadre di calciatori dilettanti o perfino professionisti.

L'autonomia di base di cui gode un termostato serve ad apprezzare quanto sia ampia la classe dei sistemi che sono autonomi nel senso dell'alternativa *i)*: essa comprende sistemi artificiali che difficilmente si direbbero intelligenti, a meno che non si faccia un uso talmente inflattivo della

parola «intelligente» da consentirne l'applicazione anche ai termostati. L'esempio del termostato serve anche a confrontare nel contesto della questione etica l'impostazione *i)* al problema dell'autonomia con *ii)-iii)*. Se queste ultime impostazioni sono utilizzate per imporre condizioni restrittive su tutta la classe dei sistemi autonomi – piuttosto che per identificarne sottoclassi particolarmente interessanti – allora si rischia di escludere sistemi che, come il termostato, svolgono un compito modesto quanto si voglia, ma pur sempre operando in piena autonomia dagli esseri umani<sup>3</sup>. L'esclusione è comprensibile se si vuole avvicinare l'autonomia dei sistemi IA al possesso di una qualche forma di intelligenza generale. Ma non lo è, se l'interesse per i sistemi autonomi è motivato dalla questione etica: vi sono armi autonome – come vedremo nel § 4 – che sollevano importanti interrogativi morali pur avendo capacità percettive e cognitive relativamente poco sviluppate rispetto allo stato dell'arte dell'IA. In definitiva, la questione etica spinge a non adottare impostazioni troppo restrittive a proposito del problema di che cosa sia una macchina autonoma.

### *3. Veicoli autonomi e conflitti morali*

Precetti morali incompatibili tra loro possono talvolta originarsi da cornici teoriche diverse nell'ambito dell'etica normativa. Consideriamo qui i conflitti che si originano dalle cornici consequenzialistica e deontologica nel caso della progettazione e dell'uso dei veicoli senza conducente.

Si ritiene che le automobili autonome saranno in grado di rispettare le norme per la sicurezza stradale con cura maggiore di un conducente umano, senza peraltro essere soggette a stanchezza, distrazione e altre condizioni psicofisiche temporaneamente invalidanti. Si tratta di un'aspettativa di grande rilevanza per le teorie consequenzialistiche in etica normativa, e in particolare per il cosiddetto «consequenzialismo delle regole». Quest'ultimo prescrive che le regole di comportamento si scelgano in funzione delle loro conseguenze attese per il benessere aggregato della popolazione di riferimento. Da tale prospettiva, permettere la circolazione di automobili autonome potrebbe presto configurarsi come una scelta da preferire a quella di proibirne l'uso.

---

<sup>3</sup> Nel manuale più diffuso di intelligenza artificiale (Stuart J. Russell e Peter Norvig, *Artificial Intelligence. A Modern Approach*. Upper Saddle River (NJ), Prentice Hall, 2010, p. 39), si identifica l'autonomia di un'agente proprio con una capacità non posseduta dai tradizionali termostati non adattivi: quella di apprendere al fine di compensare precedenti conoscenze parziali o errate.

Il consequenzialismo delle regole si articola generalmente come una teoria neutrale rispetto ai singoli: il benessere di questo o di quel particolare essere umano non ha un peso speciale nella determinazione dei benefici complessivi che sono attesi dall'adozione di una regola. L'assunzione di neutralità è rilevante per chiarire le implicazioni del consequenzialismo delle regole in uno scenario che presenta, in forma adatta ai veicoli autonomi, il cosiddetto problema morale del carrello ferroviario<sup>4</sup>. Un'automobile autonoma si trova improvvisamente davanti tre persone che hanno iniziato ad attraversare la strada. In base alla situazione, il sistema di controllo del veicolo stima di avere solo due possibilità: investire i pedoni che si trovano sulla sua traiettoria o sterzare bruscamente, portando il veicolo fuori strada e mettendo a serio repentaglio la vita di Andrea, suo unico passeggero e proprietario. Supponiamo che la casa produttrice dell'automobile autonoma abbia optato per una politica di controllo del veicolo coerente con l'assunzione consequenzialistica di neutralità. In tal caso, il sistema di controllo del veicolo sceglierà l'opzione di non investire i tre pedoni, senza mostrare particolari riguardi per Andrea.

Andrea potrà condividere o non condividere la politica di controllo del veicolo, poiché la neutralità rispetto all'insieme degli agenti non è un'assunzione condivisa da tutte le teorie etico-normative. Infatti, una cornice teorica deontologica non impegna, in generale, all'obbligo di salvare qualcuno anche a costo di mettere a serio repentaglio la propria vita. Si potrà lodare un atto del genere come gesto eroico, qualificandolo eventualmente come un atto supererogatorio, senza però identificarlo con un obbligo di colui che lo compie.

Dopo aver riflettuto sul dilemma del carrello, Andrea chiede alla casa produttrice della sua automobile senza conducente di adattarne il sistema di controllo alle sue opzioni morali di carattere deontologico. Andrea si è confrontato sulla questione con amici, conoscenti e vicini. Tra questi c'è Emma, che ha due figli piccoli. Emma sa che l'automobile autonoma di Andrea investirà i bambini che inseguono sbadatamente la palla sulla strada, se l'unica alternativa è quella di indirizzare la vettura a velocità sostenuta contro un albero. Emma ritiene che in circostanze analoghe un altro conducente umano, e perfino lo stesso Andrea a dispetto delle sue convinzioni morali, potrebbe decidere di risparmiare i bambini a scapito della propria incolumità. Emma vuole che i bambini del quartiere abbiano ancora questa possibilità, e perciò non vuole che nel quartiere circolino veicoli autonomi adattati alle preferenze di proprietari deontologicamente orientati

---

<sup>4</sup> Per una discussione di altri conflitti morali che possono interessare i veicoli autonomi si veda il saggio di Patrick Lin, *Why Ethics Matters for Autonomous Cars*, in *Autonomous Fahren*, a cura di Markus Maurer et al., Berlino, Springer Verlag, 2015, pp. 69–85. Disponibile in rete all'indirizzo [http://link.springer.com/chapter/10.1007/978-3-662-45854-9\\_4](http://link.springer.com/chapter/10.1007/978-3-662-45854-9_4).

come Andrea. In base a questo desiderio, che ella considera moralmente giustificato, Emma organizza un comitato di quartiere contro la circolazione di questi veicoli autonomi e raccoglie firme per una petizione in tal senso.

In conclusione, la riflessione etica evidenzia conflitti morali che si ripercuotono sulle scelte di progettazione dei veicoli autonomi e che possono a loro volta generare conflitti tra gruppi sociali caratterizzati da diverse opzioni morali. Per affrontare una situazione come quella che contrappone Andrea ed Emma, si potrebbe limitare il dominio di applicazione di una regola consequenzialistica a contesti d'uso nei quali non emergano richieste deontologiche incompatibili. Seguendo questa strategia generale si potrebbe, per esempio, adottare una politica di segregazione per i veicoli autonomi su strade a essi riservate<sup>5</sup>, imponendo distanze interveicolari e altre misure di sicurezza tali da escludere l'insorgere di conflitti con le istanze morali di Andrea.

#### 4. *Le armi autonome tra etica deontologica e consequenzialistica*

Consideriamo ora un analogo problema di convergenza tra etica deontologica e consequenzialistica, che riguarda il caso attualmente molto dibattuto dello sviluppo e dell'uso di armi autonome. Il Department of Defense del Governo federale degli Stati Uniti descrive le armi autonome (d'ora in poi AA) come armi che siano in grado, una volta attivate, di «selezionare e attaccare gli obiettivi senza ulteriori interventi da parte di operatori umani»<sup>6</sup>. Questa condizione applica ai compiti di selezione e attacco di un obiettivo militare la prima impostazione, descritta nel § 2, relativa al problema dell'autonomia di una macchina. Essa è soddisfatta anche da alcuni sistemi relativamente semplici, come la sentinella robotica SGR-A1. Progettata presso un'università sudcoreana, prodotta dalla ditta Samsung e dispiegata dalle forze armate della Corea del Sud, la sentinella robotica sorveglia la zona demilitarizzata di confine con la Corea del Nord. Se attivata in modalità autonoma, la sentinella individua gli intrusi nella zona demilitarizzata e decide se aprire il fuoco contro di essi, indipendentemente dal «via» di un operatore umano. Poiché nessuno ha il permesso di accedere alla zona demilitarizzata, il compito percettivo della sentinella

---

<sup>5</sup> Si veda Filippo Santoni de Sio, *Ethics and self-driving cars. A White paper on responsible innovation in automated driving systems*, Rapporto interno della Delft University of Technology, 2016.

<sup>6</sup> Si veda *Directive 3000.09: Autonomy in Weapons Systems*, US Department of Defense, 2012, pp. 13–14.

robotica si riduce alla discriminazione tra esseri umani, animali, chiome di alberi scosse dal vento e altri oggetti presenti in quell'area non urbana alla quale nessuno ha il permesso di accedere. Ma la sentinella robotica non sa risolvere altri problemi percettivi che, in generale, sono di massima importanza nei conflitti armati. Non sa distinguere tra un fuggiasco dalla Corea del Nord e chi avanza mostrando comportamenti potenzialmente ostili; tra un belligerante attivo e un nemico che è fuori combattimento perché ferito, in stato d'incoscienza o perché si è arreso con gesti difformi dallo stereotipo delle mani alzate in posizione frontale ed eretta rispetto al nemico.

Una delle ragioni avanzate per proibire le AA fa leva sulle limitazioni delle tecnologie di cui si dispone attualmente per costruire sistemi percettivi e cognitivi artificiali, le quali non consentono di sviluppare AA in grado di distinguere tra belligeranti e civili altrettanto bene di un soldato competente e coscienzioso. Più in generale, le tecnologie attuali non consentono a una macchina di rispettare, in una grande varietà di scenari bellici, tutte le regole dello *jus in bello* contemplate dal diritto umanitario in guerra (quali le Convenzioni di Ginevra e i loro protocolli aggiuntivi). Si argomenta che per questo motivo le AA debbano essere proibite almeno fino a quando le relative sfide tecnologiche siano state affrontate e superate con successo.

Questo argomento per la proibizione delle AA si colloca in una cornice etica deontologica, poiché fa un riferimento cruciale a diritti inviolabili delle persone e ad obblighi corrispondenti sanciti dal diritto umanitario in guerra, mentre si considerano irrilevanti altri benefici che potrebbero derivare dall'impiego di un'AA. La debolezza principale di questo argomento consiste nella sua dipendenza sia dallo stato attuale di sviluppo delle tecnologie e dei sistemi dell'IA (che in futuro potrebbero portare a sistemi rispettosi dello *jus in bello*) sia dal tipo di contesto bellico (poiché non tutti gli scenari operativi potrebbero porre problemi discriminativi irrisolti dall'IA).

Altri argomenti deontologici contro le AA non dipendono dallo stato di sviluppo della tecnologia e dal contesto operativo. Essi fanno appello sia all'interruzione della catena di responsabilità derivante da comportamenti non previsti dell'arma autonoma, sia al principio del rispetto della dignità umana. Seguendo questa seconda linea di pensiero, si argomenta che il rispetto della dignità umana ha implicazioni rilevanti per l'atto di sopprimere una vita umana in un conflitto armato. Quest'atto deve fondarsi su un giudizio umano informato, a garanzia di una valutazione del valore della vita umana in gioco e del significato della sua perdita. Si conclude che le AA non offrono questa garanzia e devono essere pertanto proibite. Lo stesso problema di violazione della dignità umana che nasce dall'uso di un'AA si può efficacemente rappresentare considerando la prospettiva delle sue vittime potenziali, le quali non si trovano nella condizione di poter fare



appello all'umanità di *qualcuno* che controlli l'arma e che abbia la responsabilità di decidere della loro sorte.

Consideriamo ora come stanno le cose dal punto di vista del consequenzialismo delle regole, che non ammette obblighi assoluti e diritti inviolabili, ma richiede piuttosto una valutazione aggregata di costi e benefici attesi per individuare la regola migliore di comportamento da scegliere. Alcuni vantaggi potrebbero derivare in futuro dall'utilizzazione di AA che risultino essere rispettose delle leggi umanitarie in guerra. Le AA potrebbero infatti contribuire a ridurre le perdite in entrambi gli schieramenti contrapposti sui campi di battaglia, e anche tra i civili, grazie a una loro maggiore precisione, al superamento delle limitazioni percettive, cognitive ed emotive dei comuni soldati, al fatto che al contrario di questi ultimi le AA non rischiano di perdere la vita. Tuttavia, da una prospettiva consequenzialistica più ampia che si proietta oltre i singoli campi di battaglia, bisogna considerare varie conseguenze negative derivanti dall'uso delle AA. Un elenco non esauriente di conseguenze negative comprende una nuova corsa alle armi tra attori statali, la proliferazione delle AA presso regimi oppressivi e organizzazioni terroristiche, il ricorso più facile ai conflitti armati dovuto all'aspettativa di un numero minore di morti nel proprio esercito, la conseguente destabilizzazione degli equilibri regionali e globali, i risultati di interazioni difficilmente prevedibili con altre AA di parte propria o avversa, l'accelerazione del ritmo della battaglia al di là di ogni possibilità di controllo umano, la possibilità di *escalation* da una guerra combattuta con armi convenzionali a un conflitto nucleare. Per quanto riguarda quest'ultimo punto, bisogna rilevare che sciame aerei di AA, più di molte altre armi convenzionali, potrebbero sferrare attacchi particolarmente distruttivi su obiettivi infrastrutturali e perfino nucleari, fino ad annullare la capacità dell'avversario di rivalersi con un attacco nucleare. In un tale scenario, la deterrenza nucleare basata sul *MAD* (*Mutually Assured Destruction*) perderebbe la sua efficacia ridando impulso alle strategie basate sull'attacco nucleare preventivo.

In definitiva, permettendo il dispiegamento delle AA potremmo attenderci in futuro perdite minori su alcuni campi di battaglia, ma allo stesso tempo aumenterebbe il rischio di avviare una nuova corsa alle armi, di destabilizzazioni politiche e militari, fino ad arrivare a una possibile neutralizzazione della deterrenza nucleare basata sul *MAD*. In base a questo ragionamento, si argomenta che i costi sopravanzano i benefici attesi per il benessere collettivo, e che la regola da adottare in una prospettiva consequenzialistica sia quella di vietare l'uso delle AA in tutti quei casi in cui sopravvenga un serio rischio per il mantenimento della pace e di destabilizzazione degli equilibri politici e militari.

## 5. La confluenza di ragioni deontologiche e consequenzialistiche per la proibizione delle AA

I sostenitori di una severa regolamentazione o della messa al bando totale delle armi autonome hanno utilizzato congiuntamente argomenti deontologici e consequenzialistici per sostenere le loro iniziative politiche e diplomatiche<sup>7</sup>. Tuttavia, se si adotta questa strategia, bisogna porre attenzione al fatto che alcune valutazioni morali deontologiche configgono con le valutazioni consequenzialistiche anche nel caso delle AA. Un uso selettivo di AA contro esseri umani è permesso in base al consequenzialismo delle regole, purché non sorgano seri rischi di destabilizzazione degli equilibri politici e militari. Usi di questo genere sono invece proibiti alla luce dell'argomento deontologico, basato sulla dignità umana. Analogamente, l'uso di AA contro infrastrutture disabitate non è proibito in base ad argomentazioni deontologiche, ma potrebbe esserlo in base a una valutazione consequenzialistica dei suoi effetti destabilizzanti. I sostenitori di iniziative politiche per la regolamentazione o la messa al bando delle AA hanno pertanto il problema di comporre le loro diverse argomentazioni etico-normative in un quadro unitario coerente. Vediamo se e come sia possibile risolvere questo problema.

Preliminarmente, si noti che gli argomenti deontologici contro le AA si incentrano sugli obblighi di alcune categorie di agenti (tipicamente gli operatori militari) e i diritti di alcune categorie di pazienti (le potenziali vittime delle AA). Inoltre, fanno discendere le proibizioni deontologiche contro l'uso delle AA dal principio del rispetto della dignità umana e da assunzioni relative al mantenimento della catena di responsabilità morali e legali degli operatori militari. Questi cardini delle argomentazioni deontologiche sono riconosciuti dalla comunità internazionale in quanto essi sono radicati profondamente nel diritto umanitario in guerra, nei trattati internazionali sulla tutela dei diritti umani, nonché nel diritto penale internazionale. Le norme che in questi vari ambiti riguardano l'uso della forza prevedono obblighi che non ammettono, se non in specifiche circostanze codificate, deroghe alla protezione dei diritti delle persone. Invece, da una prospettiva consequenzialistica, tali deroghe possono essere giustificate dal perseguimento di un beneficio aggregato maggiore per il genere umano. Pertanto le argomentazioni deontologiche sulla proibizione delle AA e gli obblighi sanciti dal diritto internazionale si rinforzano reciprocamente.

---

<sup>7</sup> Stephen Goose, *The case for banning killer robots*, «Communications of the ACM», LVIII, 12, 2015, pp. 43-45.

Questa concordanza offre il principale sostegno per adottare la seguente regola di priorità deontologica: le ragioni deontologiche contro le AA non possono essere scavalcate, *in ogni circostanza in cui esse sono applicabili*, da ragioni consequenzialistiche a favore dell'uso delle AA. Ma che cosa accade quando le ragioni deontologiche contro le AA non offrono nessuna guida morale? Una circostanza del genere si verifica quando non sono in gioco gli obblighi di alcune tipologie di agenti o i diritti non negoziabili di alcune tipologie di pazienti. Limitandosi al quadro teorico deontologico, si potrebbe assumere che in tali circostanze è permesso tutto ciò che non è deontologicamente proibito. Tuttavia l'etica consequenzialistica non rimane silente a proposito di questa zona grigia per l'etica deontologica: essa impone infatti di proibire tutte le AA che potrebbero avere effetti destabilizzanti per la pace e la sicurezza. È da notare che una tale proibizione consequenzialistica è rinforzata anche dai principi enunciati nell'art. 1 della Carta delle Nazioni Unite e da altri strumenti giuridici relativi alla pace e alla sicurezza internazionale. In sintesi, ai fini di iniziative politiche volte a introdurre una severa regolamentazione o la messa al bando totale delle AA, la regola di priorità deontologica consente di distribuire le ragioni deontologiche e consequenzialistiche su diversi domini di applicazione e di ottenere per questa via una convergenza coerente delle due cornici etiche normative, che è rinforzata anche dalle norme del diritto internazionale e dei principi che ad esso soggiacciono<sup>8</sup>.

## 6. Osservazioni conclusive

Sono state discusse due soluzioni locali ai conflitti interteorici tra etica deontologica e consequenzialistica nel contesto tecnologico delle macchine autonome. Le soluzioni proposte consentono di raccordare in modo coerente una pluralità di modelli ideali dell'etica normativa a processi decisionali concreti, nei quali si riscontrano e spesso convivono istanze deontologiche e consequenzialistiche. Le soluzioni locali proposte per i veicoli senza conducente e le AA sono rilevanti per il problema generale di fornire orientamenti morali per inquadrare ed effettuare scelte di progettazione e di utilizzazione per le macchine autonome. Analoghi problemi che richiedono una riflessione filosofica approfondita sono sollevati dagli algoritmi per il gioco in

---

<sup>8</sup> Questo modello di convergenza è stato proposto e le sue motivazioni etiche e giuridiche sono state più ampiamente discusse in Daniele Amoroso e Guglielmo Tamburrini, *The confluence of deontological and consequentialist reasons for banning autonomous weapons systems*, Preprint dell'Università di Napoli Federico II, 12 ottobre 2016.

borsa, dagli agenti autonomi *software* per la guerra cibernetica e da altre macchine autonome che l'IA consente di costruire.

*Summary. Autonomous Machinery and AI Philosophy*

Philosophical motives of interest for AI and robotic autonomous systems prominently stem from distinctive ethical concerns: in which circumstances are autonomous systems ought to be permitted or prohibited to perform tasks which have significant implications in the way of human responsibilities, moral duties or fundamental rights? Deontological and consequentialist approaches to ethical theorizing are brought to bear on these ethical issues in the context afforded by the case studies of autonomous vehicles and autonomous weapons. Local solutions to intertheoretic conflicts concerning these case studies are advanced towards the development of a more comprehensive ethical platform guiding the design and use of autonomous machinery.

Keywords: Autonomous Vehicles, Autonomous Weapons, Deontological and Consequentialist Ethics.